Machine Learning
&
Remote Sensing Imagery

James Theiler
Space and Remote Sensing Sciences
Nonproliferation and International Security
Los Alamos National Laboratory

2002 December 4

Workshop on...

Image Analysis and Understanding Data from Scientific Experiments

Purpose of this Talk

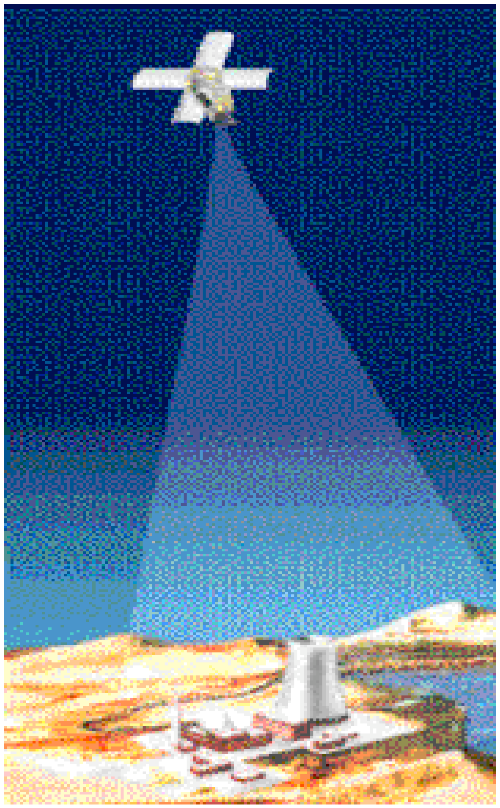
- Convince you that machine learning has a useful place in the analysis of imagery

- Argue that "the remote sensing problem" is a scientific experiment whose output data are in need of some (peace, love, &) understanding

- Note that imagery produces new challenges (a.k.a. opportunities) for machine learning

The Remote Sensing Problem

- Given satellite imagery, what's on the ground?

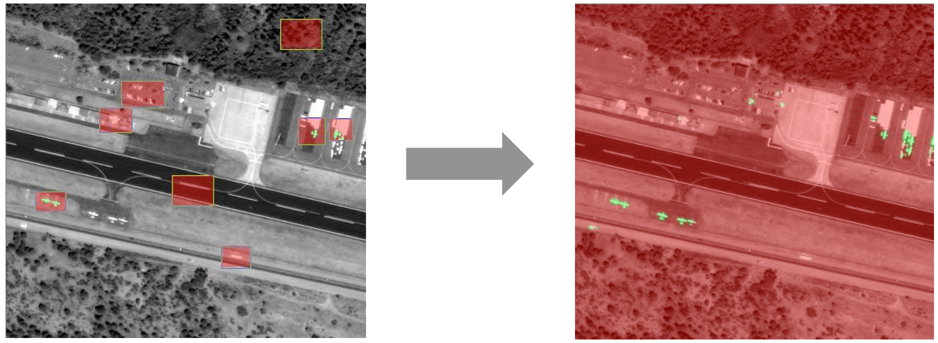


- Material identification
  - Broad Area Features (lakes, forests, *etc.*)
  - Small Targets (vehicles, runways, *etc.*)
- Plume (weak signal) detection
- Spectral *and* spatial signatures

Huge Inverse Problem

- Materials illuminated from sun, sky, reflections
- In thermal bands, direct emission from ground
- Radiative transfer through a dynamic atmosphere
- Sensor: optics, focal plane, electronics

Promise of Machine Learning



Easier to show a machine what to find…

- ML derives classification algorithms directly from examples of data

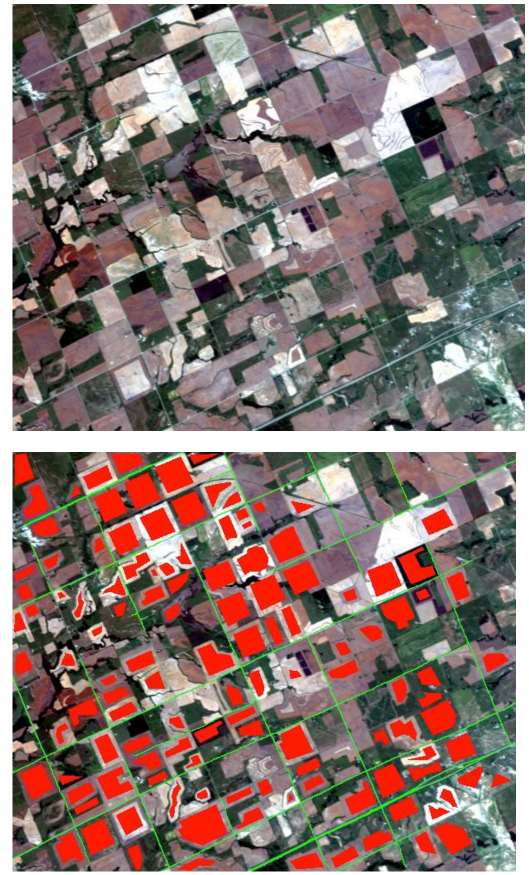
…than to tell a machine how to find it

- Requires deep understanding of problem domain
- Writing good algorithms is a specialized skill
- Design process can be slow and laborious
- Performance of algorithm can be difficult to characterize

Landcover

Urban Areas

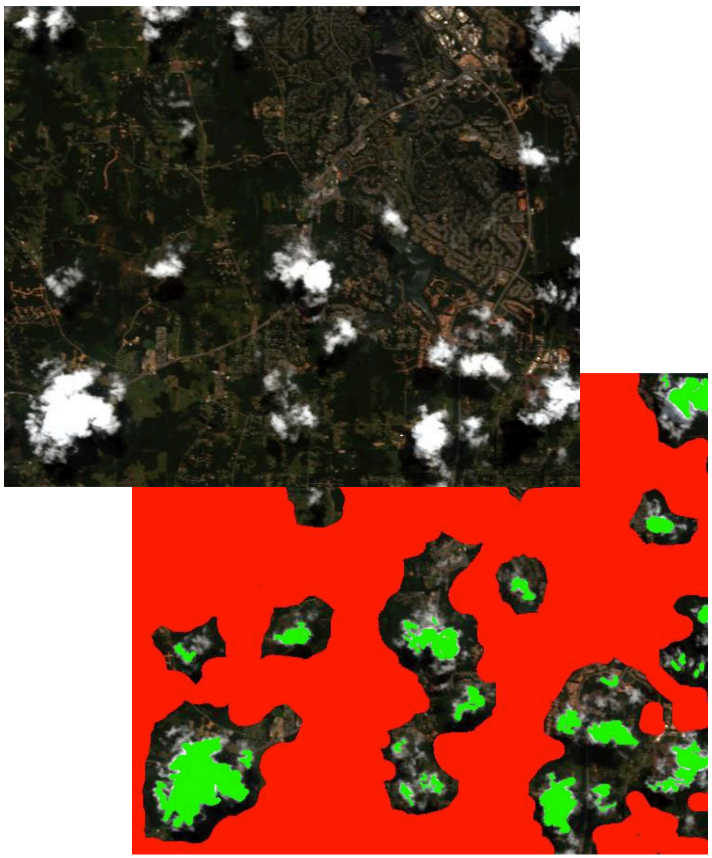
Golf Courses

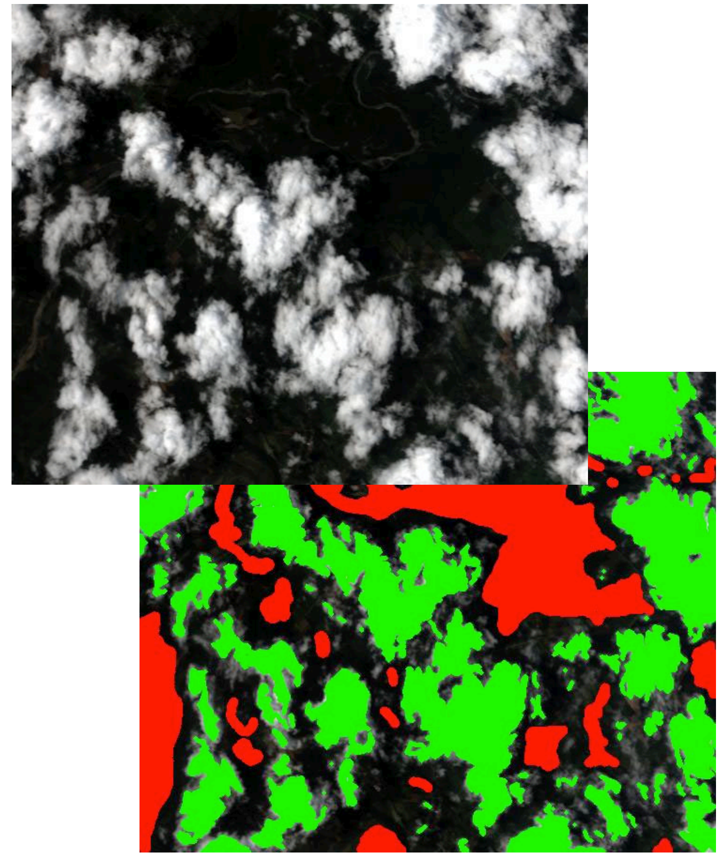


Roads

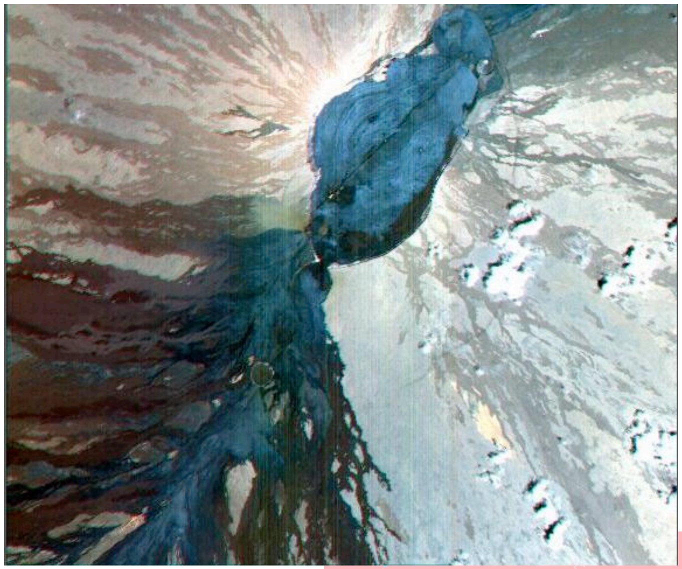
Clouds

- Signal for some; background for others
- Bright, white, cold (and dry!)

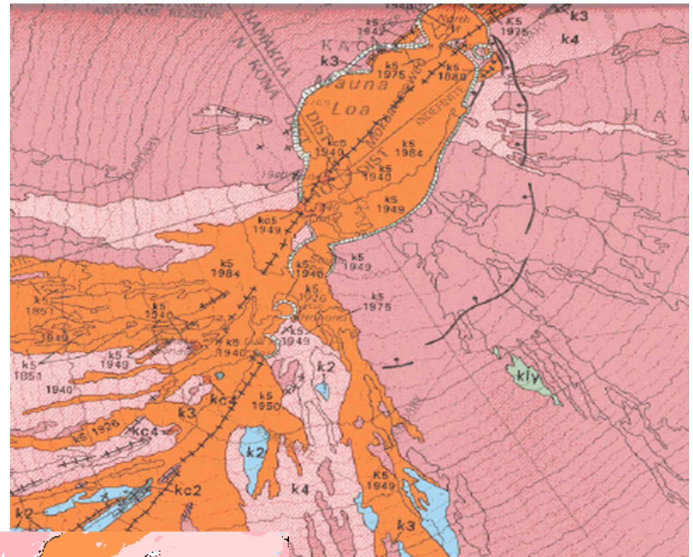




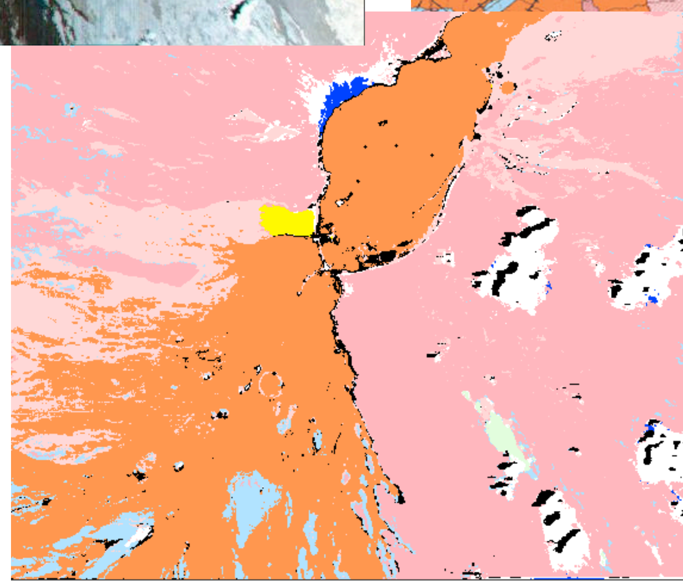
Mineral mapping on Mauna Loa



MTI Data

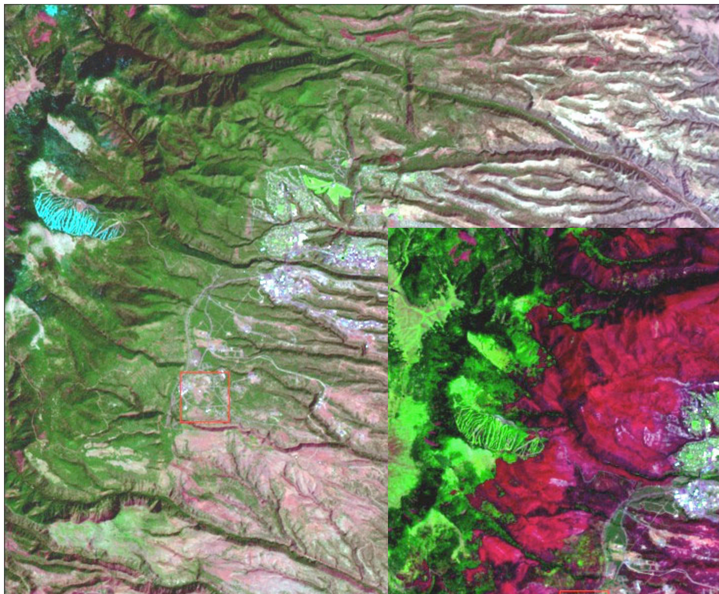


USGS Map



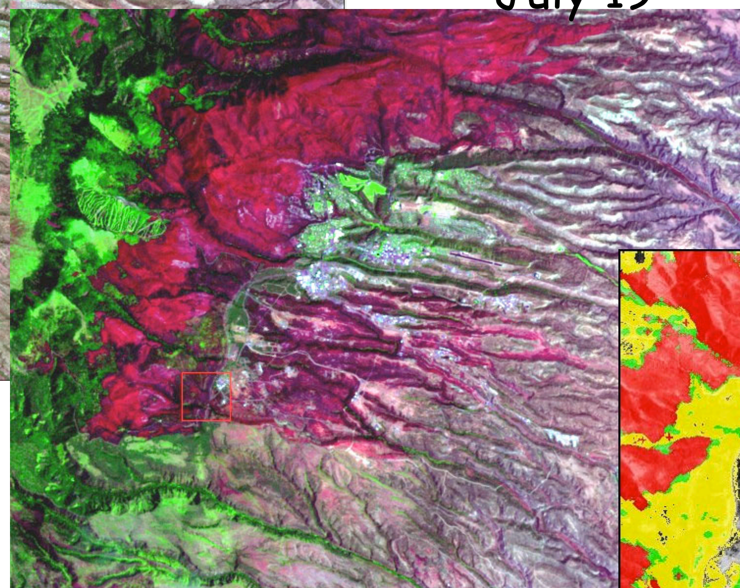
Multiclass Classification

Cerro Grande Fire: Before and After

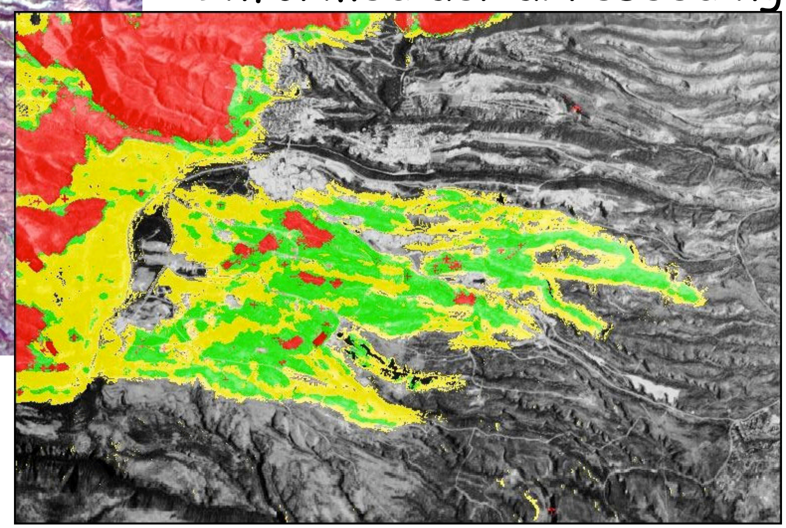


April 14

July 19

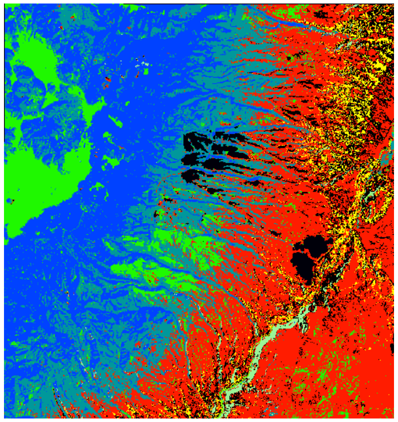


Burn Severity Map
based on Dadaelus MSI
Informed aerial reseeding

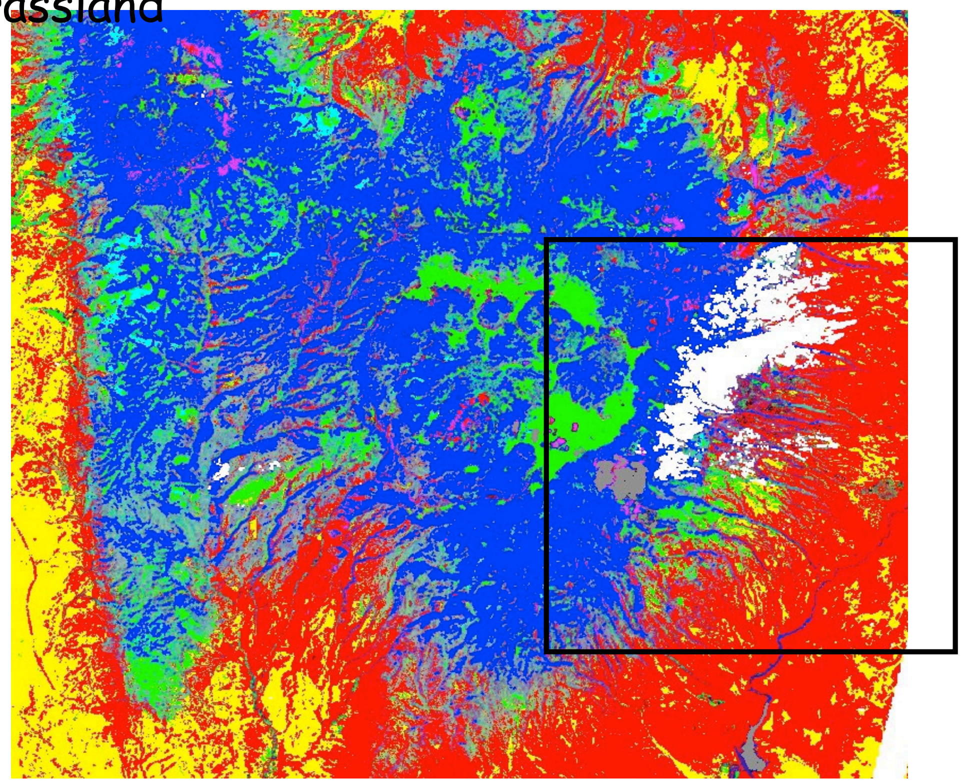


NASA Landsat 7
Multispectral Imagery
Bands 7,4,3

Land Cover Classification

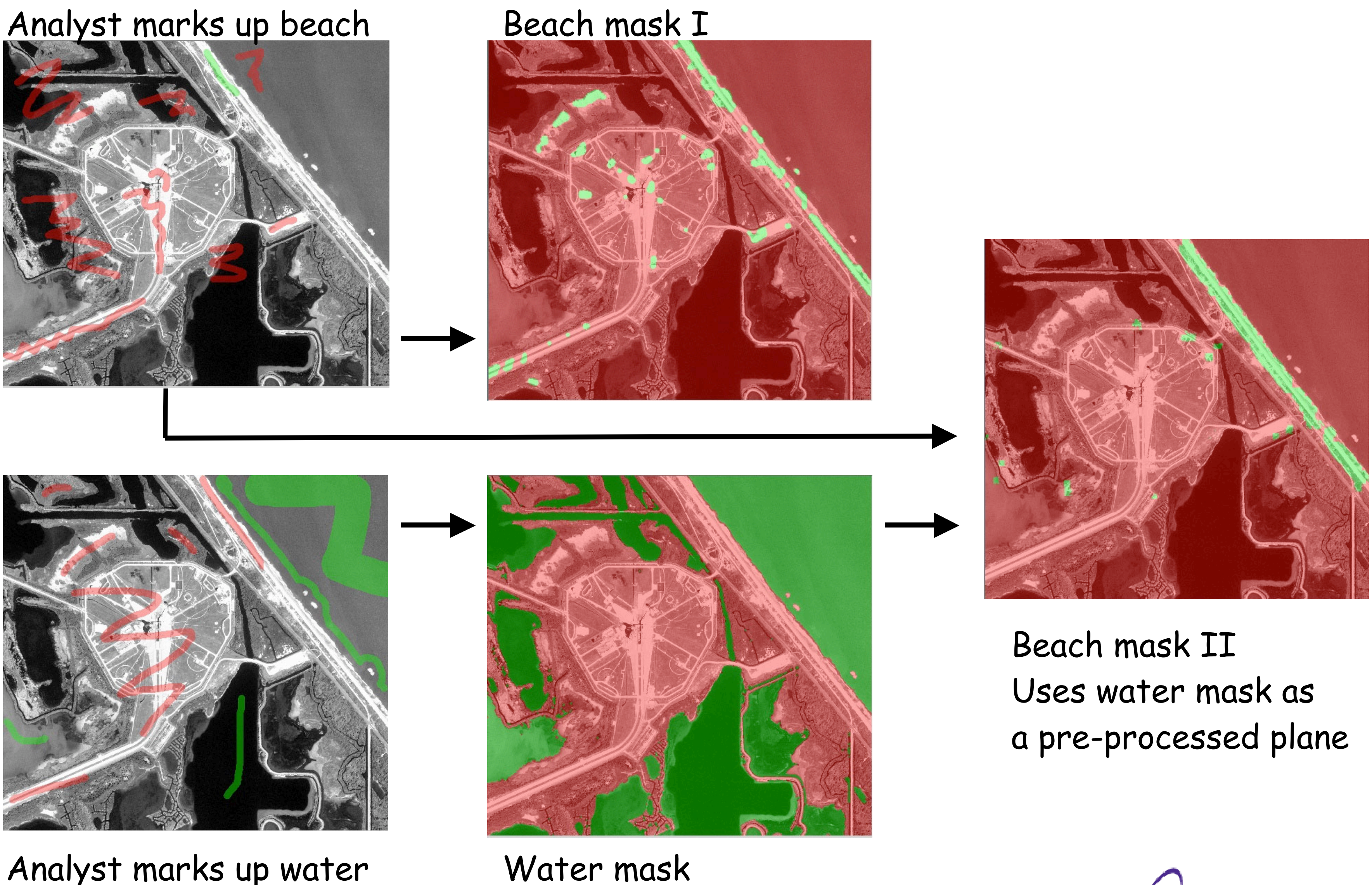


- "Official" classification map, based on ground truth from field excursions and Aug 1992, Landsat 5 TM data.
- Main Classes
  - Red: Pinon/Juniper
  - Green: Open grassland
  - Blue: Forest
- Townsites in black



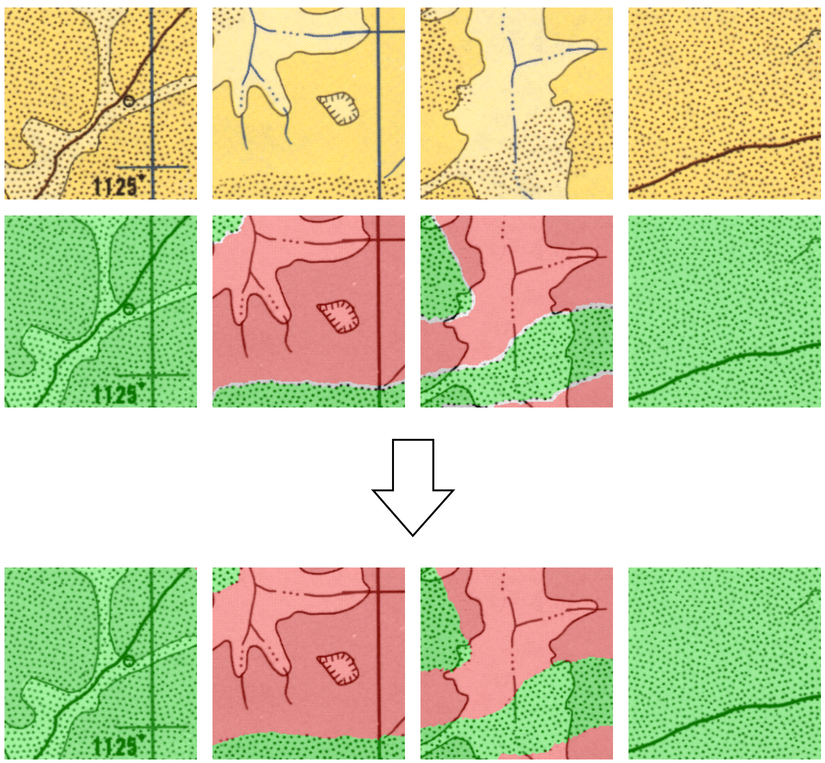
- Genie classifiation map, based on post-fire Landsat 7 ETM+ data
- Trained four classes
  - Red, Green, Blue: from official
  - White: Fire damage
- Covers much larger region
- Needed for Elk Habitat Study

Expert Assistance: use water to find beach

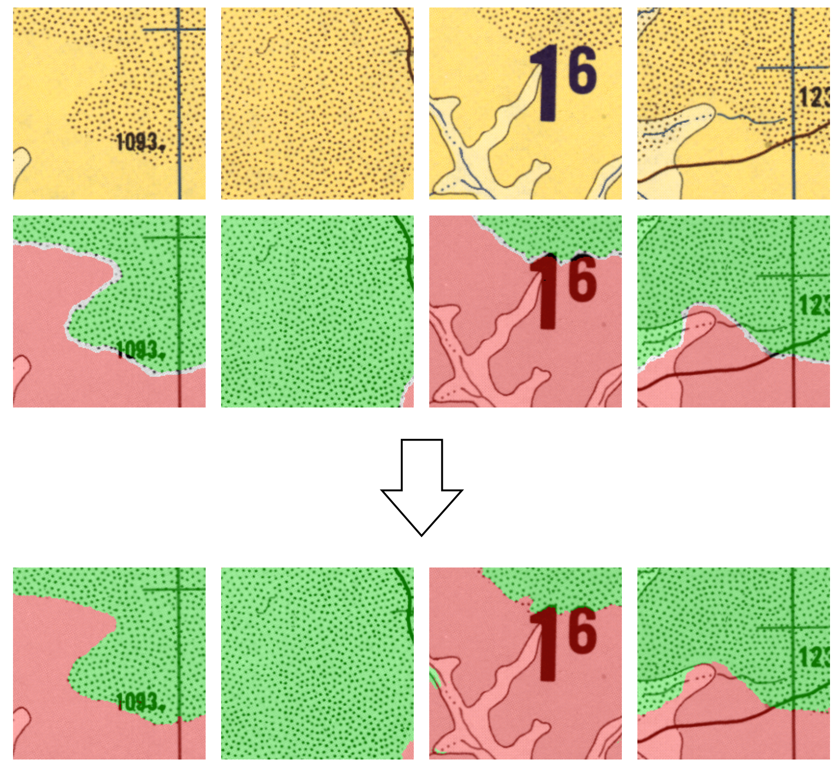


Map Vectorization

Training Data

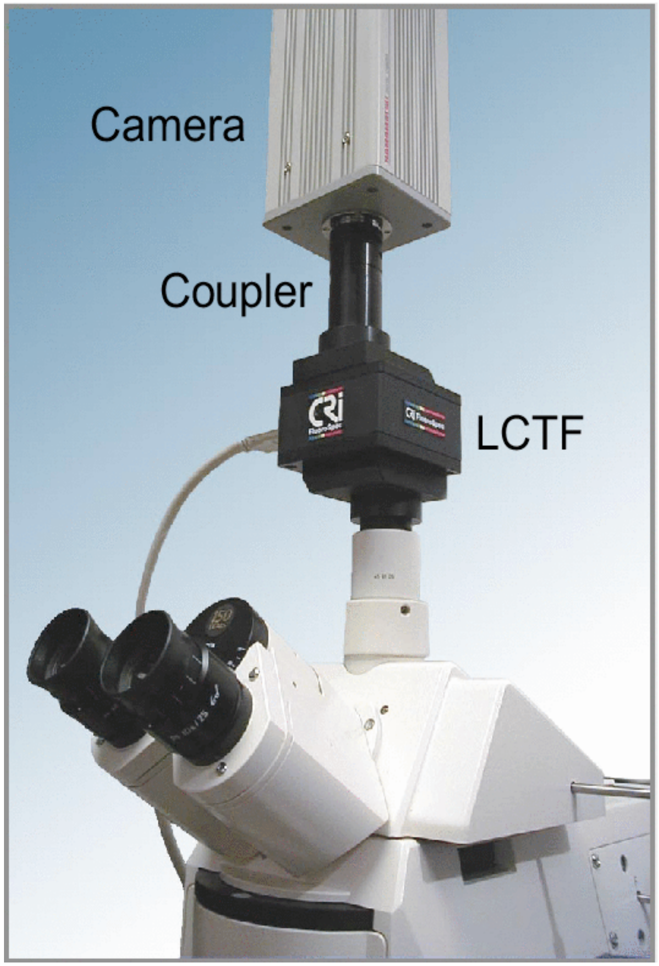


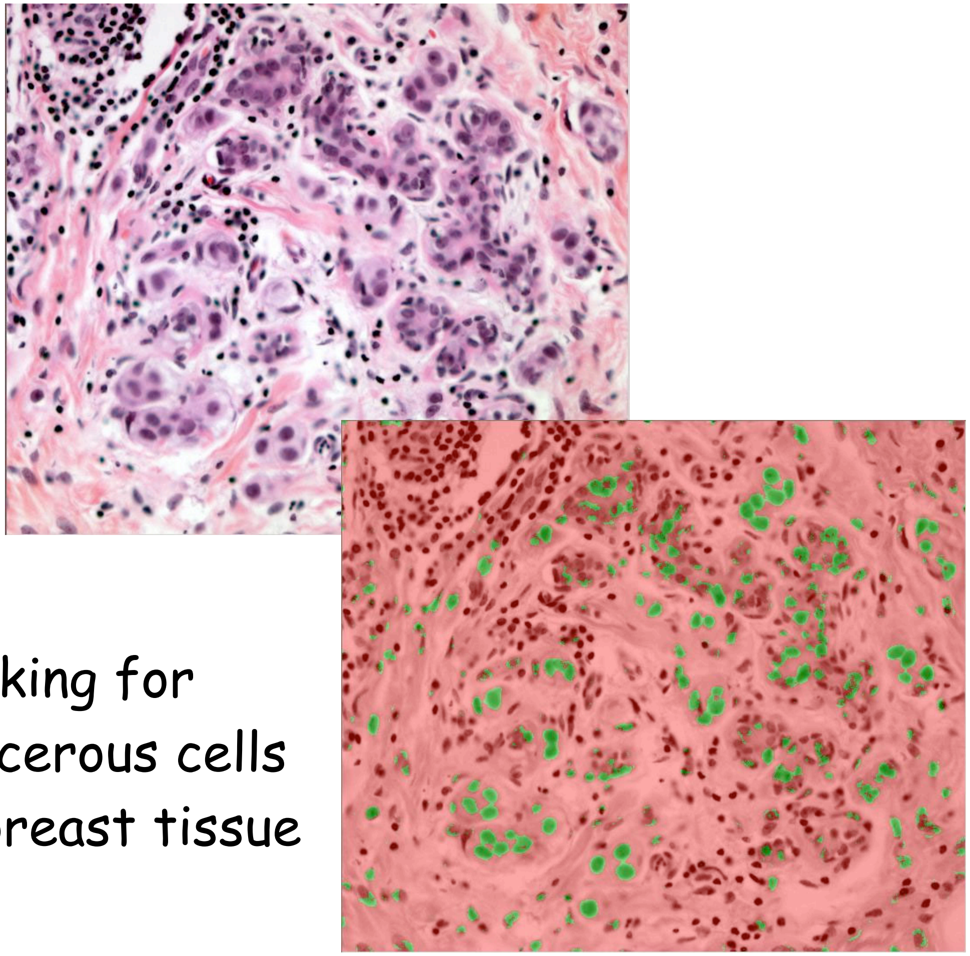
Test Data



Multispectral Microscope Imagery

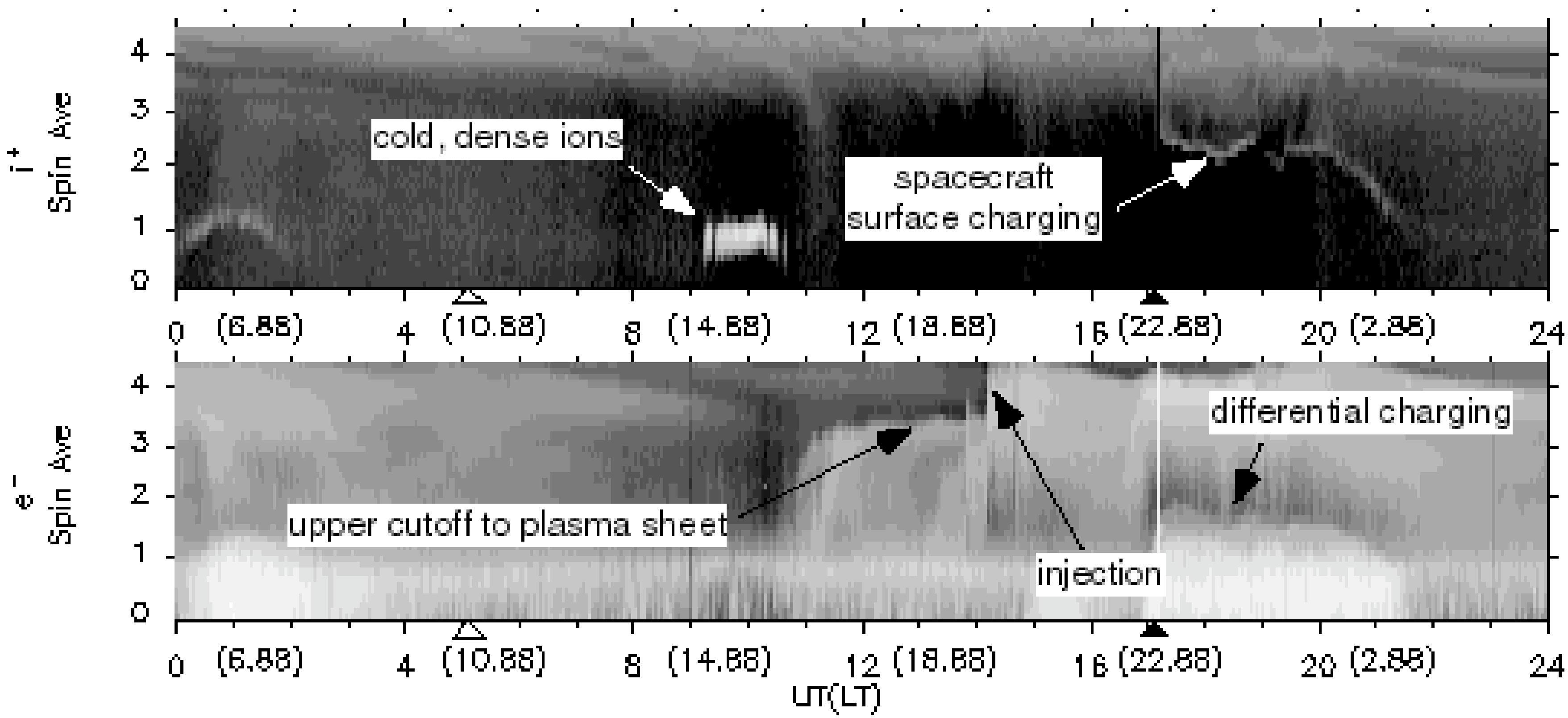
24 bands, 450 – 680 nm @ 10 nm intervals



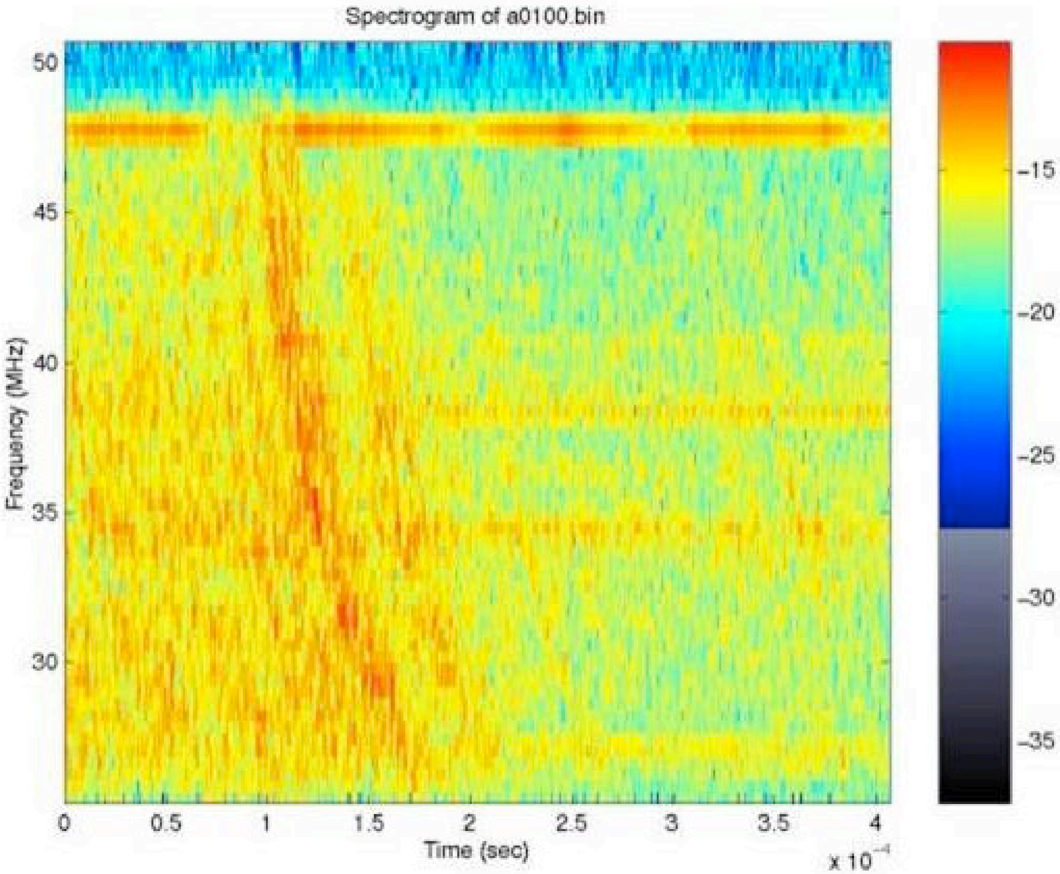


Looking for cancerous cells in breast tissue

Annotating Space Weather



Signals represented as images

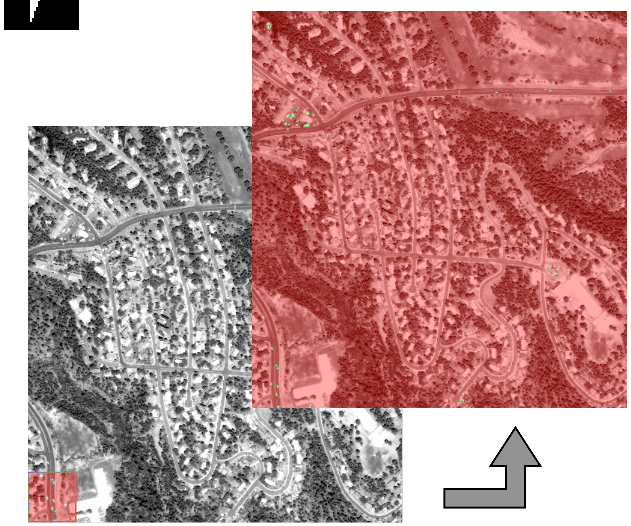


- Forte' Data
- Time-Frequency Histogram
- Translational invariance in only one direction

Target identification

Locate and characterize individual craters on the surface of Mars

ID aircraft from silhouette

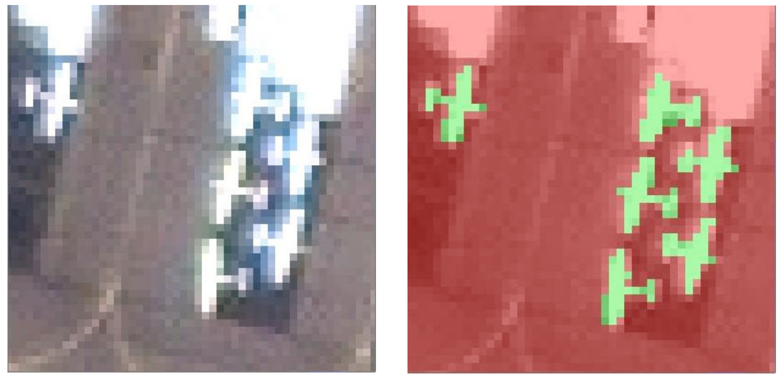


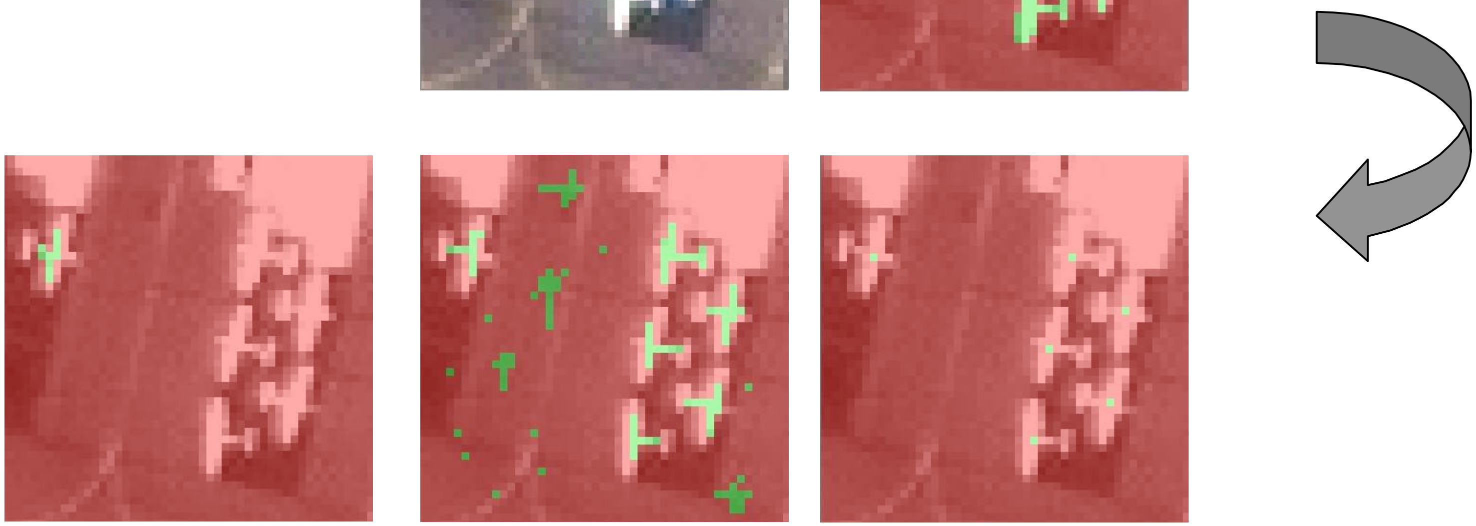
Find cars in IKONOS imagery
(How to exploit all that unlabelled data?)

Focus of Attention

Finding pixels vs. finding airplanes

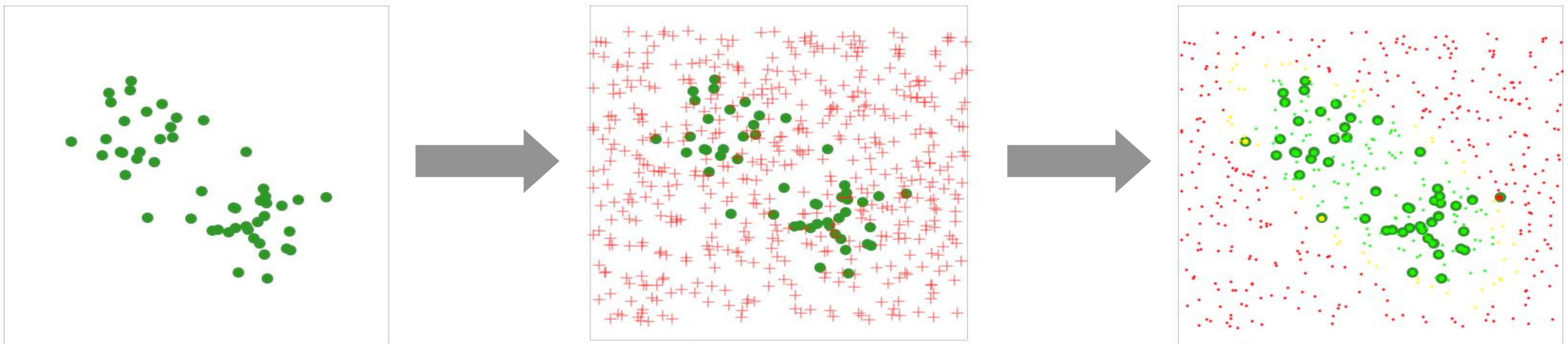
Pixel-by-pixel training data:



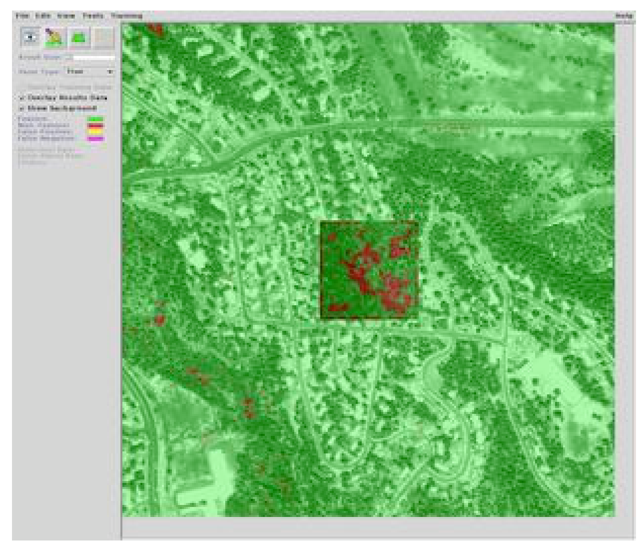


Which result is preferable?

Anomaly Detection



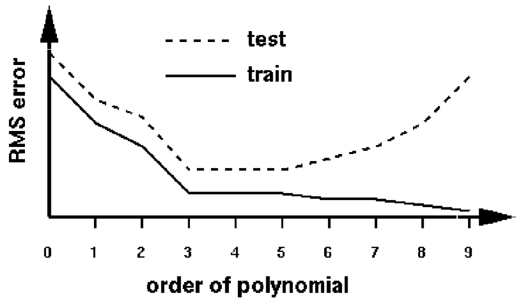
- Find the "unusual" pixels in a scene
- Recast as a two-class problem
  - Normal class exemplified by data
  - Anomalous class is uniform background
- Distinguish data from "random" using conventional ML (support vector machine)
- Exploring variations on "random"
- Applications to change detection

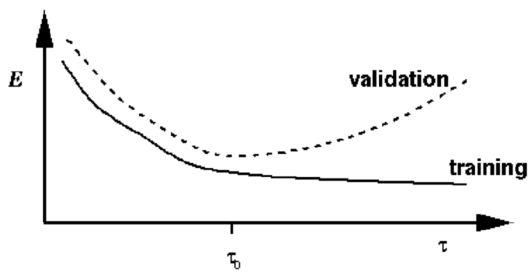


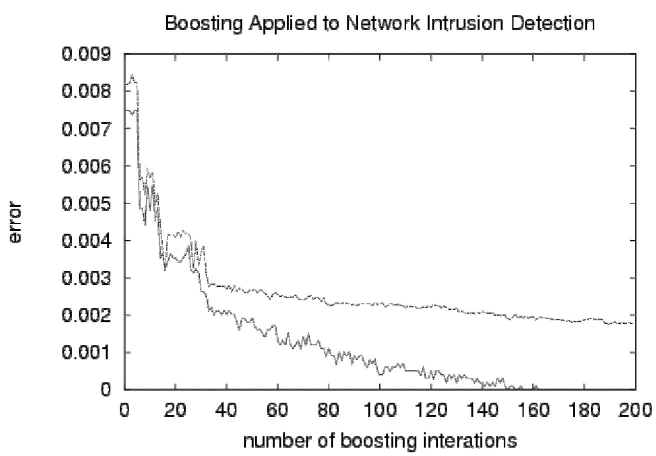
Machine Learning: fitting predictive models to data

- Given training data (x_i, y_i), $i=1,\ldots,m$
- Find a function $f(x)$ for which $y_i=f(x_i)$
  - Fit data in-sample: $y_i=f(x_i)$
  - Fit data out-of-sample: $y_i=f(x_i)$ for $i>m$.
- In practice, fits are approximate:
  - Error function: $E(f) = (1/m)\_{i}\, L(y_i, f(x_i))$
  - Squared loss: $L(y,f(x)) = (y-f(x))^2$
  - Margin based, eg: $L(y,f(x)) = exp(-yf(x)/\_)$

Flexibility vs Overfitting







- Complexity of Classifier
  - better to fit the data
  - more prone to overfitting
  - Occam's razor: use the simplest model that fits the data.
  - VC Theory: formalizes tradeoff
- Traditional ML approaches employ *ad hoc* methods to balance complexity and in-sample error
  - Cross-validation
  - Regularization
  - Limited training time
- SVM's and boosting produce seemingly "complex" classifiers without overfitting.

Why are images different?

- Pixels are not independent samples
  - Contiguity effects
  - Spatial correlations
- Focus-of-attention issues
  - Don't always care about precise pixel-wise classification